2段標本選択線形予測法による高ピッチ音声の分析

著者	三好 義昭, 大和 一晴, 柳田 益造, 角所 収
雑誌名	電子情報通信学会論文誌 A 基礎・境界
巻	70
号	8
ページ	1146-1156
発行年	1987-08-01
URL	http://hdl.handle.net/2297/3349

論　文

2段標本選択線形予測法による高ピッチ音声の分析

正　員　三好　義昭[†]　　正　員　大和　一晴[†]
正　員　柳田　益造[††*]　　正　員　角所　収[††]

Analysis of Speech Signals of Short Pitch Period by a Two-Stage Sample-Selective Linear Prediction

Yoshiaki MIYOSHI[†], Kazuharu YAMATO[†], Masuzo YANAGIDA[††*] *and* Osamu KAKUSHO[††], *Members*

あらまし　近年，音声認識などにおいて重要となる声道伝達特性の推定手法として線形予測法が広く用いられているが，通常の線形予測法はピッチ周期の影響を受け易いという欠点があるため，特に女声のような高ピッチ音声の声道伝達特性の正確な推定がしばしば困難となる．本論文では，筆者らが先に提案した標本選択線形予測法を改善することにより，高ピッチ音声の声道伝達特性の推定が通常の線形予測法より正確に行えることを示す．標本選択線形予測法は，線形予測法における残差情報に基づき線形予測モデルにより適合する音声標本のみに線形予測関係をあてはめ予測係数を推定しているため，励振源の影響を軽減できるという利点がある．ここで提案する2段標本選択線形予測法は，標本選択を更に推し進めて2段階に拡張し，かつ残差信号の大局的な特徴も考慮して，線形予測モデルに適合する音声標本を選択することにより，高ピッチ音声の声道伝達特性の推定精度向上を目指したものである．本方法の精度および有効性は，高ピッチ周期の合成音のホルマント周波数推定精度の改善，および成人女性が発声した単音節の母音部の分析例と連続音声のホルマント周波数抽出の改善により示されている．

1. まえがき

音声の伝送・認識において，声道伝達特性を正確に推定することは極めて重要であり，今日その手段として線形予測法[(1),(2)]が広く活用されている．しかしながら，音声生成系が全極型モデルで記述でき，かつ定常的であると考えられる音声区間においても，通常の線形予測法を用いて正確な声道伝達特性が得られるためには，理論的には，励振源が単一のインパルスあるいは白色雑音でなければならないが，現実にはその仮定は満たされていないので，ホルマント周波数推定に励振源の影響が生じる．特に，女性あるいは子供が発声した有声音のように基本周期の短いいわゆる高ピッチ音声の場合，ホルマント周波数推定はピッチの影響を大きく受けて正確な推定が困難となる状況がしばしば生じる．この有声音における励振源の影響を軽減する方法としては，有声音源のより実際的なモデル化[(3)]，あるいは分析窓長を1ピッチ周期以下と短くして声門閉止区間すなわち自由振動区間のみを分析対象とする方法[(4)~(6)]などがある．有声音源のモデル化は，声帯波形関数の推定および位相の問題などまだ未解決の重要な問題があり，今後の研究課題であると言える．一方，自由振動区間内分析では，声門閉止区間を正確に推定しておく必要があり，種々の方法[(7),(8)]が検討されてはいるが，自然音声の声門閉止区間を正確に推定するのは一般に困難で，特に女声のような高ピッチ音声の場合には，声門閉止区間の推定はより困難となる．本論文では，励振源の影響を受けない正確な分析には，基本的には自由振動区間を対象とした処理が現時点では最良であるとの立場から，従来の自由振動区間内分析における上記の難点のない分析法としての標本選択線形予測法の改良を示す．

筆者らは，先に，線形予測分析における予測残差に基づき励振源を含まないと見なせる部分の音声標本を被予測標本として選択的に使用する標本選択線形予測

† 姫路工業大学電子工学科，姫路市
Himeji Institute of Technology, Himeji-shi, 671-22 Japan
†† 大阪大学産業科学研究所，茨木市
Institute of Scientific and Industrial Research, Tsaka University, Ibaraki-shi, 567 Japan
* 現在，郵政省電波研究所

分析を提案した[9]．この手法は線形予測分析に一般逆行列を導入し，Givens変換に基づく逐次計算法を用いることによって各標本値を処理するごとにその選択的利用が効率良く行える利点があった．しかしながら，それは処理時間節減の目的から，各標本値を処理するごとにその標本値を使用するか否かを決定していたため，予測残差の大局的な特徴に基づく選択処理が行えない欠点があった．本論文では標本の選択処理をフレーム単位で行うことにより，予測残差の大局的な特徴をも考慮して標本の選択を行い，かつこの処理を2段階行うことにより，従来の方式よりも被予測標本としてより妥当な標本の選択が行える2段標本選択線形予測法を提案する．以下，2.において，本方法の基本的な考え方を示し，3.で合成音のシミュレーションにより本方法のホルマント周波数推定精度の改善を示す．4.では，本方法を実際に成人女性が発声した単音節の母音部の分析例ならびに連続音声のホルマント周波数抽出に適用した例を示し，通常の線形予測法では正確な分析が比較的困難であった女声のような高ピッチ音声の分析に対して本手法が特に有効であることを示す．

2. 2段標本選択線形予測法

周知のように音声の線形予測法は，音声生成系を

$$s_n=\sum_{k=1}^{p}\alpha_k s_{n-k}+u_n \qquad (1)$$

但し，

s_n：音声波の第 n 標本値

u_n：励振源の第 n 標本値

α_k：予測係数

なる全極型モデルで記述できるものと仮定し，予測係数 $\alpha_k(k=1,2,3,\cdots,p)$ の推定値 $\hat{\alpha}_k$ を予測誤差の2乗平均最小の条件より，

$$\sum_{k=1}^{p}\hat{\alpha}_k\phi_{ik}=\phi_{i0},\quad i=1,2,\cdots,p \qquad (2)$$

但し，

$$\phi_{ik}=\sum_{n=p+1}^{N}s_{n-i}s_{n-k}$$

なる正規方程式の解として求めるものである（共分散法）．

ところで，式(2)の正規方程式は

$$\boldsymbol{S}^T\boldsymbol{S}\hat{\boldsymbol{\alpha}}=\boldsymbol{S}^T\boldsymbol{s} \qquad (3)$$

但し，

$$\boldsymbol{S}=\begin{bmatrix} s_p & s_{p-1} & s_{p-2} & \cdots & s_1 \\ s_{p+1} & s_p & s_{p-1} & \cdots & s_2 \\ s_{p+2} & s_{p+1} & s_p & \cdots & s_3 \\ \vdots & \vdots & \vdots & & \vdots \\ s_{N-1} & s_{N-2} & s_{N-3} & \cdots & s_{N-p} \end{bmatrix}$$

$$\hat{\boldsymbol{\alpha}}=(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3,\cdots,\hat{\alpha}_p)^T$$

$$\boldsymbol{s}=(s_{p+1},s_{p+2},s_{p+3},\cdots,s_N)^T$$

と書ける．一方，式(1)を $n=p+1,p+2,\cdots,N$ について一括して行列の形式で表現すると，

$$\boldsymbol{S}\boldsymbol{\alpha}+\boldsymbol{u}=\boldsymbol{s} \qquad (4)$$

但し，

$$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\cdots,\alpha_p)^T$$

$$\boldsymbol{u}=(u_{p+1},u_{p+2},u_{p+3},\cdots,u_N)^T$$

と記述できる．式(4)の両辺に左から $\boldsymbol{S}^T$ を掛けると，

$$\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{\alpha}+\boldsymbol{S}^T\boldsymbol{u}=\boldsymbol{S}^T\boldsymbol{s} \qquad (5)$$

となるので，式(3)，(5)より

$$\boldsymbol{S}^T\boldsymbol{S}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha})=\boldsymbol{S}^T\boldsymbol{u} \qquad (6)$$

となる．すなわち，

$$\boldsymbol{S}^T\boldsymbol{u}=\boldsymbol{0} \qquad (7)$$

となれば $\hat{\boldsymbol{\alpha}}=\boldsymbol{\alpha}$ となり正確な予測係数が得られることになる．有声音の場合，観測された音声波 $\boldsymbol{s}$ に適当な前処理を施すことによって，励振源 $\boldsymbol{u}$ は既周期的なパルス列とみなすことができるので，成人男性の有声音のようにピッチ周期が比較的長ければ式(7)が近似的に成立するため，通常の線形予測法により予測係数を精度よく推定できるが，成人女性あるいは子供の有声音のようにピッチ周期が短いいわゆる高ピッチ音声の場合には式(7)が近似的にも成立しなくなり，推定精度が悪くなる危険性がある．

ところで，有声音の場合，分析窓長を1ピッチ周期以下とし，いわゆる声門閉止区間のみを分析対象とすれば $\boldsymbol{u}=\boldsymbol{0}$ となるため，予測係数を精度よく推定できると言えるが，自然音声の声門閉止区間を正確に推定するのは一般に困難であり，特に女声のような高ピッチ音声の場合には，それはより困難となるだけでなく，声門閉止区間が推定できたとしてもその区間長が極端に短くなるため，予測係数の個数と予測式（式(8)参照）の個数が同程度にしかならず分析フレーム内にわずかでも励振があることによる影響ならびに雑音の影響を過敏に受け分析結果のフレーム間連続性に問題が生じやすいと言える．

ところで，式(3)より，通常の線形予測法は予測係数の推定値 $\hat{\alpha}_k$ を

2　分析窓長 T_a=20～30 ms の通常の線形予測分析を行い予測係数を求める．

3　得られた予測係数に基づき残差信号 e_n を計算する．但し，分析フレーム内での残差信号の絶対値の最大値を与える値（符号を含む）で正規化する．すなわち，残差信号は本質的に双極性であるため，この正規化により予測信号の全体的な極性を正極性に正規化することになる．

4　残差がしきい値 θ 以下となる音声標本 $\{s_{n_1}, s_{n_2}, s_{n_3}, \cdots, s_{n_M}\}$ を選定．但し，2回目の標本選択処理では残差がしきい値 θ 以上となる音声標本の手前 N_0 個を除く（図1参照）．

5　選定された音声標本を被予測標本とする予測式を連立させ，その最小2乗解（式(9)の解）として，予測係数を求める．

$$S_M^I S_M \tilde{\alpha} = S_M^I s_M \qquad (9)$$

但し，

$$S_M = \begin{bmatrix} s_{n_1-1} & s_{n_1-2} & s_{n_1-3} & \cdots & s_{n_1-p} \\ s_{n_2-1} & s_{n_2-2} & s_{n_2-3} & \cdots & s_{n_2-p} \\ s_{n_3-1} & s_{n_3-2} & s_{n_3-3} & \cdots & s_{n_3-p} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{n_M-1} & s_{n_M-2} & s_{n_M-3} & \cdots & s_{n_M-p} \end{bmatrix}$$

$$s_M = (s'_{n_1}, s'_{n_2}, s'_{n_3}, \cdots, s_{n_M})^T$$

6　ステップ3に戻り3～5の処理を再度行う．

3.　合成音による分析精度の検討

標本化周波数10 kHz, 励振源：ピッチ周期3.8 msのRosenberg波[10]（図1(a)参照），ホルマント周波数：表1，放射特性：6 dB/octとして作成した合成5母音を用いて，本方法によるホルマント周波数推定精度の改善を明らかにする．

合成母音/o/における各部の波形ならびに標本選択の例（図1(c)および(d)の下段｜印）を図1に示す．但し，前処理として一階差分，分析次数 p=12, 分析窓長 T_a=25.6 ms とした．図1(c)は通常の線形予測分析による残差波形，図1(d)は図1(c)の残差波形に基づいて前章の手順により一度標本選択線形予測分析（しきい値 θ=0.2）を行った場合の残差波形である．両者の残差波形を比較すると，標本選択線形予測分析による残差波形の方が通常の線形予測分析による残差波形よりパルス列状に近くなっていると言える．本方法はこの特徴を積極的に利用したものである．すなわち，従来の標本選択処理による標本選択の例（図1(c)の下段｜印，但し，しきい値 θ=0.2 とし，残差の絶対値

表1　合声音のホルマント周波数

(Hz)

	F_1	F_2	F_3	F_4	F_5
/a/	813	1313	2688	3438	4438
/i/	375	2188	2938	3438	4438
/u/	375	1063	2188	3438	4438
/e/	438	1813	2688	3438	4438
/o/	438	1063	2688	3438	4438

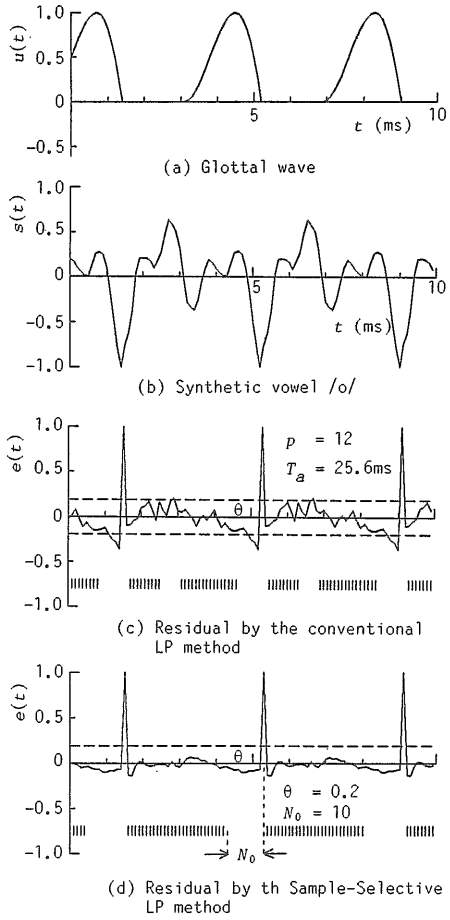


図1　合成母音/o/における残差波形および標本点選択の例（各残差波形の下段｜印は選択された標本点を示す）

Fig. 1　An example of speech waveform and the corresponding residual signals with marks of sample selection (｜) for a synthetic vowel /o/.

が θ 以上となる点およびその手前3点，後1点を除去）から明らかなように，従来の標本選択処理でも各声門閉止点およびその手前の声門開口部付近に対応する音声標本が被予測標本から除かれてはいるが，正確な声道伝達特性推定に必要な声門閉止部付近の音声標本も被予測標本から除かれる（今の場合，t=2.8 ms お

よび6.6 ms付近)ことがあると言える．それに対して本方法では，一度標本選択線形予測分析した後のよりパルス列状に近くなっている図1(d)の残差波形に基づいて標本を選択しているため（しきい値 $\theta=0.2$, 予測残差がしきい値 θ 以上となる手前の除去標本点数 $N_0=10$ とした場合，図1(d)の下段｜印となる)，各声門閉止点およびその手前の声門開口部付近に対応する音声標本のみが被予測標本から除かれ，より適切な標本が選択されていると言える．なお今の場合，しきい値 $\theta \geqq 0.3$ とすれば，従来の標本選択処理でも本方法とほぼ同等の標本選択が行えると言えるが，自然音声の場合，通常の線形予測分析による残差が図1(c)程度のパルス列状にはならず，また母音定常部でも分析フレーム内の各ピッチごとの残差のピークレベルにかなりの差が生じることがあるため，従来の標本選択処理では適切なしきい値 θ を設定することが困難となる．この点に関しては4.で具体的に述べる．

なお，ここでは標本の選択処理を2段階で留めているが，標本選択処理により予測残差が大きくなる音声標本すなわち線形予測モデルに適合しない音声標本が被予測標本から除かれていくので，原理的にはもっと多段階行ってもなんら問題はないと言える．しかしながら，3段階以上行っても合成音ならびに自然音声とも顕著な改善がみられなかったので，標本の選択処理の簡素化を考慮して，2段階に固定した．

3.1 しきい値 θ の検討

式(10)で定義される合成5母音の第1～第3ホルマント周波数推定誤差の平均値 E のしきい値 θ 依存性を図2に示す．但し，前処理として1階差分後，分析次数 $p=12$, 分析窓長 $T_a=25.6$ ms とし，フレームシフト 0.2 ms で1周期に渡って分析した場合の平均値である．そして○印：本方法で $N_0=10$ とした場合の結果，△印：従来の標本選択線形予測分析の結果である．また，通常の線形予測分析の誤差を図中…にて示す．

$$E=\frac{1}{15}\sum_{j=1}^{5}\sum_{i=1}^{3}|\hat{F}_{ij}-F_{ij}|/F_{ij} \qquad (10)$$

但し，

F_{ij}：第 j 母音の第 i ホルマント周波数

$\hat{F}_{ij}$：第 j 母音の第 i ホルマント周波数推定値

図2より，ホルマント周波数推定誤差が通常の線形予測分析ではピッチの影響によりり5.3％と大きかったものが，従来の標本選択線形予測分析により2.4％程度に改善し，更に本方法により0.9％と大幅に改善されていることがわかる．そして，従来の標本選択線形予測法の誤差は $0.3 \leqq \theta < 1.0$ においてはしきい値 θ に依存せず一定であるが，$\theta < 0.3$ においてしきい値 θ により変動しているのに対し，本方法の誤差は $0.1 \leqq \theta < 1.0$ においてしきい値 θ に依存しないことが分かる．ここで用いた合成音では，$\theta \geqq 0.3$ とすれば，従来の標本選択線形予測法においても残差信号がしきい値 θ 以上となるのは各ピッチごとの実効的な励振点のみとなるため（図1(c)参照)，誤差は $0.3 \leqq \theta < 1.0$ においてしきい値 θ に依存せず一定となる．したがって，$\theta < 0.3$ においてはじめて標本選択処理を2回行う効果が得られていると言える．なお，$\theta \geqq 0.3$ において本方法の誤差が従来の標本選択線形予測分析の誤差より改善しているのは本方法における除去標本点数 N_0 の効果によるものである．すなわち，各ピッチごとの実効的な励振点付近のみを除去するよりも，残差レベルのいかんにかかわらず実効的な励振点とその手前10

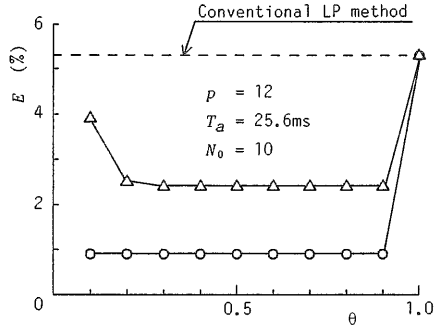


図2 ホルマント周波数推定誤差のしきい値 θ 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法

Fig. 2 The θ (threshold)-dependency of the formant estimation error E.
○: Two-stage SSLP method.
△: Prototype SSLP method.

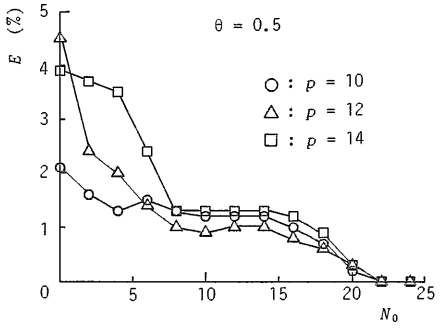


図3 ホルマント周波数推定誤差の N_0 依存性

Fig. 3 The N_0-dependency of the formant estimation error E.

点程度を除去すれば，ホルマント周波数推定誤差が大幅に改善すると言える．

3.2 N_0 の効果

本方法におけるホルマント周波数推定誤差 E の N_0 依存性を図3に示す．但し，しきい値 $\theta=0.5$ とし，○印：$p=10$, △印：$p=12$, □印：$p=14$ とした場合の結果である．図3より，$N_0>0$ すなわち残差がしきい値 θ 以上となる音声標本値の手前 N_0 個を残差レベルのいかんにかかわらず被予測値から除外することにより，ホルマント周波数推定誤差が減少し，$0<N_0\leqq 8$ においては推定誤差の改善度合が分析次数により異なるが，$N_0\geqq 22$ で分析次数にかかわらず推定誤差が零となることがわかる．

有声音の場合，音声波に適当な前処理をほどこせば，励振源は概周期的なパルス列とみなすことができると言えるが，これは近似的に言えることであり実際には完全なパルス列とはならない．本シミュレーションでは励振源として図1(a)に示すRosenberg波を用いた．したがって，音声波を2階差分すれば，励振源は各声門閉止時点でほぼパルス状とはなるが(今の場合，放射特性として1階差分を用い，分析の前処理として1階差分を行っているので，実質的に励振源を2階差分したことになる)，各声門開口区間では零とはならない．ところで，前章で明らかにしたようにホルマント周波数推定精度は式(7)の成立度合に依存している．したがって，$N_0>0$ とすることにより各声門開口区間に対応する音声標本が残差レベルの大きさのいかんにかかわらず被予測標本から除外されるためホルマント周波数推定精度が改善されることになる．本シミュレーションに用いた励振波形の声門開口区間は2.2 msであるので，$N_0\geqq 22$ とすれば，各声門開口区間に対応する音声標本は被予測標本から完全に除外されることになり，いわゆる声門閉止区間内分析(式(7)が成立)となるため，ホルマント周波数推定誤差は零となる．したがって，合成音の分析に際しては N_0 をできるだけ大きくして，声門閉止区間内分析に近づければ良いと言えるが，自然音声，特に高ピッチ音声の場合，N_0 をあまり大きくすると予測式の個数が少なくなり分析結果の安定性に問題が生じるので N_0 をあまり大きくとることはできない．このことを考慮すれば，図3の結果より $N_0=10$ 程度が適当と思われる．

合成5母音について従来の方法と本方法($\theta=0.5$, $N_0=10$)のホルマント周波数推定誤差の比較を表2に示す．表2より，母音/a/において本方法の誤差が従来の標本選択線形予測法の誤差より若干悪くなっているが，他の母音に関してはいずれも誤差が更に改善し，特に高ピッチにおいてピッチ周波数と第1ホルマント周波数が接近し，ピッチの影響を大きく受けると思われる母音/i/および/u/の改善が著しいと言える．

3.3 分析次数の検討

ホルマント周波数推定誤差 E の分析次数 p 依存性を図4に示す．但し，○印：本方法($\theta=0.5$, $N_0=10$)，△印：従来の標本選択線形予測法($\theta=0.5$)，×印：通常の線形予測法（共分散法）で，分析次数 $p=12$, 分析窓長 $T_a=25.6$ ms とした場合の結果である．

図4より，従来の標本選択線形予測法の誤差は分析

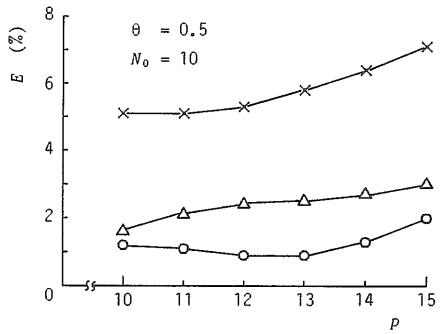


図4　ホルマント周波数推定誤差の分析次数 p 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

Fig. 4　The p(analysis order)-dependency of the formant estimation error E.
○：Two-stage SSLP method.
△：Prototype SSLP method.
×：Conventional LP method.

表2　ホルマント周波数推定誤差の比較　(%)

母音	/a/	/i/	/u/	/e/	/o/
通常の線形予測法	1.5	6.7	5.2	6.8	6.1
従来の標本選択線形予測法	0.3	3.6	3.5	2.1	2.3
本方法	0.6	0.4	0.4	1.6	1.7

次数と共に若干単調に増大していたのが，本方法により分析次数依存性が改善し，かつ，ホルマント周波数推定誤差はいずれの分析次数においても通常の線形予測法による誤差よりも大幅に小さいと言える．

3.4 ピッチ周期に関する頑健性の検討

ホルマント周波数推定誤差 E のピッチ周期 T_s 依存性を図5に示す．但し，合成音は励振源のピッチ周期のみを3.0 msから5.0 msまで変化させたものであり，他の合成条件（励振源の声門開口比など）および分析条件は図4と同じである．なお各印の意味も図4と同じである．

図5より，ピッチ周期が4.2 ms以上では従来の標本選択線形予測法による誤差も本方法による誤差もほぼ同じであるが，従来の標本選択線形予測法はピッチ周期が4.2 ms以下になるとホルマント周波数推定精度が徐々に悪くなっていたのが，本方法によりピッチ周期が4.0 ms以下の誤差が大幅に改善されていると言える．なお，通常の線形予測法による誤差はピッチ周期が4.6 ms以下になると急激に大きくなっていたのが，ピッチ周期が4.0 ms以下になると誤差は平均的には減少する傾向がみられる．これはピッチ周期とホルマント周波数の相対位置関係により，誤差が小さくなったものと考えられる．ホルマント周波数がピッチ周波数の高調波間の1/4あるいは3/4付近に位置する場合，通常の線形予測法では正確なホルマント周波数推定が一般に困難となるが[11]，ここで用いた合声音の各ホルマント周波数はピッチ周期が4.0 ms程度のときにその状態となるため，それよりもピッチ周期が短い場合にはかえって推定誤差が小さくなるという傾向になっている．

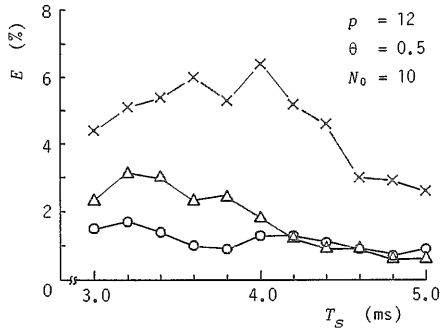


図5 ホルマント周波数推定誤差のピッチ周期 T_s 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

Fig. 5 The T_s (pitch period)-dependency of the formant estimation error E.
○: Two-stage SSLP method.
△: Prototype SSLP method.
×: Conventional LP method.

4. 自然音声への適用例

標本選択処理を成人女性が発声した単音節/bo/の母音部の/o/（ピッチ：約3.2 ms）に適用した場合の残差波形および標本点選択の例を図6に示す．但し，標本化周波数10 kHz，前処理として1階差分後，分析次数 $p=12$，分析窓長 $T_a=25.6$ ms，$\theta=0.5$，$N_0=10$ とした場合の例である．図6（b）は通常の線形予測分析による残差波形，図6（c）は図6（b）の残差波形に基づいて2.の手順により一度標本選択線形予測分析（しきい値 $\theta=0.5$）を行った場合の残差波形である．なお，従来の標本選択法および本方法により被予測標本点として選択された選択時点をそれぞれ図6（b）および（c）の下段｜印にて示す．図6（b）の通常の線形予測分析による残差波形より明らかなように，自然音声の

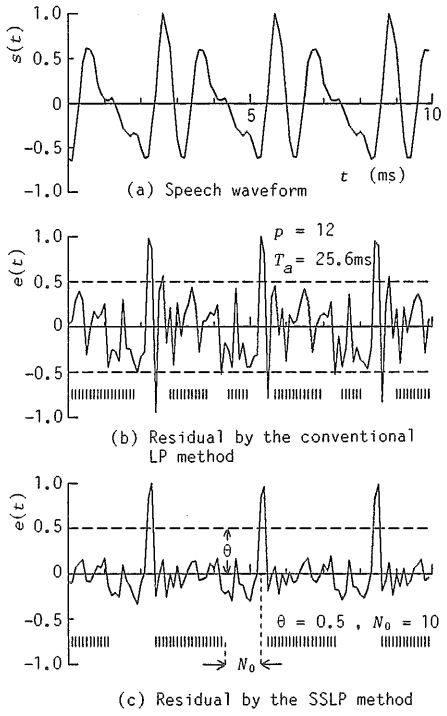
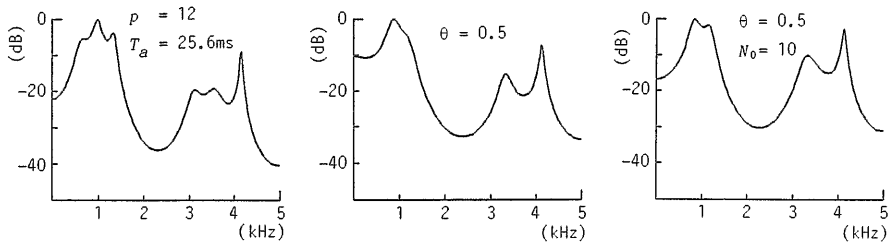


図6 自然音声（女声：母音/o/）における残差波形および標本点選択の例
（各残差波形の下段｜印は選択された標本点を示す）

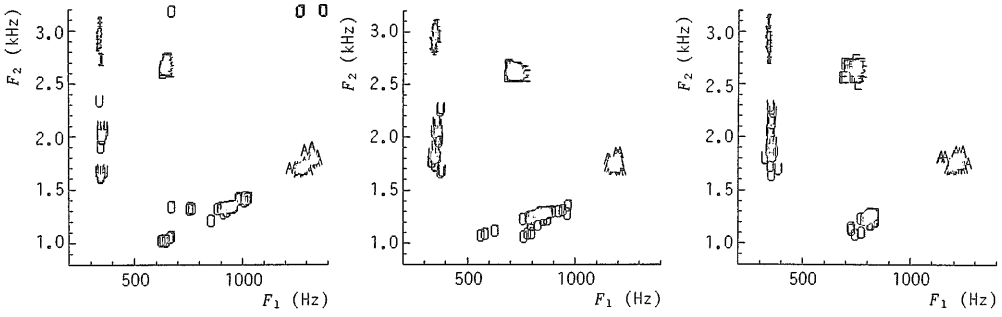
Fig. 6 An example of speech waveform and the corresponding residual signals with marks of sample selection (|) for a natural vowel /o/ uttered by a female.



(a) Conventional LP method (b) Prototype SSLP method (c) Two-stage SSLP method

図7 スペクトル包絡の比較
(女声：母音/o/)

Fgi. 7 A comparison example of the spectral envelopes of a natural vowel /o/ uttered by a female.



(a) Conventional LP method (b) Prototype SSLP method (c) Two-stage SSLP method

図8 ホルマント空間における女声の5母音分布

Fig. 8 Distribution of the natural vowels uttered by a female on the F_1-F_2 plane.

場合，通常の線形予測分析による残差は合成音のようなパルス列状（図1（c）参照）とはならないため，従来の標本選択法においてしきい値 $\theta=0.5$ としても適切な標本の選択が行えていない．それに対して，2. の手順により1度標本選択線形予測分析（しきい値 $\theta=0.5$）を行った場合の残差（図6（c））は通常の線形予測法による残差（図6（b））よりもよりパルス列状となっており，この残差に基づいて再度標本選択処理（$\theta=0.5, N_0=10$）を行えばより適切な標本点選択ができることが分かる．今の場合，従来の標本選択法でもしきい値 $\theta\geqq0.6$ とすれば良いと言えるが，一方自然音声では母音定常部でも分析フレーム内の各ピッチごとの残差のピークレベルにかなりの差が生じることがあるため，θ をあまり大きく設定できない場合がある．すなわち，従来の標本選択法では実際に女声のような高ピッチ音声を分析する場合，しきい値 θ を入力音声に応じていかに適応的に設定するかを解決する必要がある．これに対して本手法では適切なしきい値 θ の許容範囲が広いため（図6の場合，$0.2\leqq\theta<1.0$），このしきい値設定問題は回避できたと言える．

図6の場合のスペクトル包絡の比較を図7に示す．但し，（a）：通常の線形予測法によるスペクトル包絡，（b）：従来の標本選択線形予測法（$\theta=0.5$）によるスペクトル包絡，（c）：本方法（$\theta=0.5, N_0=10$）によるスペクトル包絡である．図7より，次のことが言える．通常の線形予測法では1kHz付近および3.2kHz付近に近接する3個および2個の明確な極が存在しいずれがホルマントか判断しがたい．従来の標本選択線形予測法により3.2kHz付近のスペクトル包絡が改善され第3ホルマントが明確となっているが，第2ホルマントが不明確である．本方法ではこれらの点がすべて改善されており第1～第3ホルマントが明確となっている．

F_1-F_2 平面上の5母音の分布の比較を図8に示す．但し，音声資料は成人女性1名が発声した単音節（70種）の母音定常部各3フレーム，分析条件は図7の場合と同じであり，得られた極のうちバンド幅の小さいものをホルマントとみなした．

図8より，次のことが言える．通常の線形予測法では母音/o/の分布のバラツキが大きく，かつ単なるバンド幅の情報のみではホルマントを誤推定する場合があり（$(F_1, F_2)=(600\,\text{Hz}, 3.2\,\text{kHz})$ および $(1300\,\text{Hz}, 3.2\,\text{kHz})$ 付近の計6フレーム），また母音/u/の分布が2クラスに分かれている．従来の標本選択線形予測法ではバンド幅の情報のみでもホルマントを誤推定することがなく，また母音/u/の分布が改善しているが，母音/o/の分布のバラツキがあまり改善されていない．本方法ではこれらの点がいずれも改善されており，通常の線形予測法と比較し特に母音/o/の分布の改善が著しいと言える．

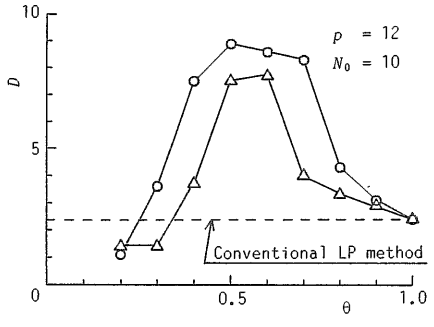


図9　ホルマント空間における分離度 D のしきい値 θ 依存性
○：2段標本選択線形予測法
△：従来の標本選択線形予測法

Fig. 9　The θ (threshold)-dependency of the separation degree D on the F_1-F_2 plane.
○：Two-stage SSLP method.
△：Prototype SSLP method.

ホルマント空間上における分布の良さを評価するために，F_1-F_2 平面上における5母音の分布の類内分散と類間分散に基づいた分離度 D を式(11)で定義し，そのしきい値 θ 依存性を図9に示す．

$$D=\sqrt{\frac{\sum_{k=1}^{5}(m_k-m)^T(m_k-m)}{\frac{1}{N}\sum_{k=1}^{5}\sum_{x\in X_k}(x-m_k)^T(x-m_k)}} \tag{11}$$

但し，

$x=(F_1, F_2)^T$

$m_k=\frac{1}{N}\sum_{x\in X_k} x$

$m=\frac{1}{5}\sum_{k=1}^{5} m_k$

N：資料数/クラス（今の場合，$N=42$）

図9より，従来の標本選択線形予測法においても，しきい値 θ が $0.5\leqq\theta\leqq0.6$ の範囲であれば分離度 D は大きくなるが，最適なしきい値 θ の範囲が狭いと言える．これに対して，本方法の分離度は $0.4\leqq\theta\leqq0.7$ においてしきい値 θ にほとんど依存せず大きな分離度

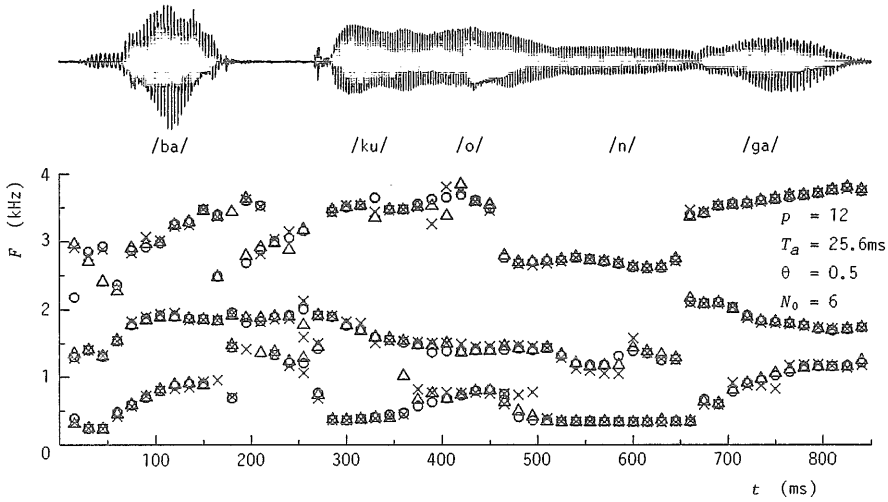


図10　連続音声（女声：/bakuonga/）のホルマント周波数抽出例
○：2段標本選択線形予測法
△：従来の標本選択線形予測法
×：通常の線形予測法

Fig. 10　An example of formant trajectories for the continuous speech /bakuonga/ uttered by a female.
○：Two-stage SSLP method.
△：Prototype SSLP method.
×：Conventional LP method.

が得られており，本方法の有効性が示されていると言える．

成人女性が発声した連続音声「爆音が」のホルマント周波数抽出に適用した例を図10に示す．但し，前処理として一階差分後，分析次数 $p=12$, 分析窓長 $T_a=25.6$ ms, フレーム間隔 15 ms とし，○印：本方法 ($\theta=0.5$, $N_0=6^{\dagger}$), △印：従来の標本選択線形予測法 ($\theta=0.5$), ×印：通常の線形予測法による結果である．なお，ホルマント周波数は各フレームごとに 200 Hz～4 000 Hz に得られた極のうちバンド幅の小さいものを第1～第3ホルマントとして抽出した．

図10より，次のことが言える．通常の線形予測法ではホルマント周波数の時間的変化に不自然な不連続が生じている ($t=330$ ms 付近の F_2, $t=380$ ms 付近の F_1 および F_3, $t=500$ ms 付近の F_1, $t=600$ ms 付近の F_2 ならびに $t=750$ ms 付近の F_1)．また，$t=465$ ms の所では 700 Hz 付近に近接する明確な極が存在するため第3ホルマントが抽出できていない．従来の標本選択線形予測法により，これらの不連続性がかなり改善されてはいるが，$t=380$ ms 付近の第3ホルマントの時間的変化がまだ不連続であり，また $t=360$ ms の所で第1ホルマントに誤抽出が生じている．これに対して，本方法ではこれらの不連続性が更に改善されており，本方法の有効性が示されていると言える．

5. むすび

残差情報を参照することによって線形予測モデルに適合する音声標本点を選択する標本選択線形予測法において，標本の選択処理を予測残差の大局的な特徴を考慮して行い，かつこの処理を2段階行う2段標本選択線形予測法の有効性を検討した．その結果，本方法は従来の標本選択線形予測法よりも被予測標本としてより妥当な標本の選択が行えることが明らかとなった．そして，本方法は通常の線形予測法ではピッチの影響により正確な分析がしばしば困難であった高ピッチ音声の分析に特に有効であることが，合成音によるホルマント周波数推定精度の改善，自然音声のスペクトル包絡の改善と抽出したホルマント周波数分布の改善ならびに連続音声のホルマント周波数抽出の改善により明らかとなった．

なお，本方法を自然音声に適用するにあたり，本方法のパラメータである除去標本点数 N_0 を固定としたが，2. で述べたように N_0 は声門開口区間の音声標本をできるだけ被予測標本から除くために導入したパラメータであるので N_0 の最適値は声門開口区間すなわちピッチ周期に依存する量であると言える．したがって，特に連続音声に適用する場合には，N_0 はピッチ周期に応じて適応的に変化させることが望ましいと言えるが，この点に関しては今後の課題である．

† 本連続音声はピッチ周期が 2.6～3.5 ms に渡って変化しているため，ここではこの値を用いた．

謝辞 本研究に関し有益な御助言を頂いた阪大産研溝口理一郎助教授ならびに御討議頂いた阪大産研電子機器部門の各位に深く感謝する．

文　献

(1) 板倉，斉藤：“統計的手法による音声スペクトル密度とホルマント周波数の推定”，信学論(A)，**53-A**, 1, pp. 35-42 (昭45-01).

(2) B. S. Atal and S. L. Hanauer：“Speech analysis and synthesis by linear prediction of the speech”, J. Acoust. Soc. Amer., **50**, 2, pp. 637-655 (1971).

(3) M. Ljungquist, 藤崎：“線形予測分析にもとづく声帯音源・声道パラメータの同時推定法”，音響学会音声研資，**S85-21** (昭60-06).

(4) S. Chandra and W. C. Lin：“Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis”, IEEE Trans. Acoust., Speech & Signal process., **ASSP-22**, pp. 403-415 (1974).

(5) 河原，栃内，永田：“小区間の線形予測分析とその誤差評価：，日本音響学会誌，**33**, 9, pp. 470-479 (昭52-09).

(6) K. Steiglitz and B. Dickinson：“The use of time-domain selection for improved linear prediction”, IEEE Trans. Acoust., Speech & Signal process., **ASSP-25**, pp. 34-39 (1977).

(7) H. W. Strube：“Determination of the instant of glottal closure from the speech wave”, J. Acoust. Soc. Am., **56**, pp. 1625-1629 (1974).

(8) T. V. Ananthapadmanbha and B. Yegnanarayama：“Epoch extraction of voiced speech”, IEEE Trans. Acoust., speech & Signal process., **ASSP-23**, pp. 562-570 (1975).

(9) 溝口，柳田，谷口，角所：“一般逆行列を用いた音声の選択的線形予測分析” 信学論(A)，**J66-A**, 1, pp. 56-63 (昭58-01).

(10) A. E. Rosenberg：“Effect of Glottal Pulse Shape on the Quality of Natural Vowels”, J. Acoust. Soc. Am., **49**, pp. 583-590 (1971).

(11) 藤崎，佐藤：“音声のホルマント抽出の諸方式の比較検討”，音響学会音声研資，**S47-1** (昭49-05).

(昭61年12月22日受付，62年3月19日受付)

三好　義昭

昭42姫路工大・電気卒．同年同大電子工学科助手．音声の分析および認識，ディジタル信号処理などの研究に従事．日本音響学会会員．

大和　一晴

昭29姫路工大・電気卒．同形同大助手．講師，助教授を経て，昭46電子教授．工博．現在，多値論理，画像処理，音声認識および信頼性に関する研究に従事．電気学会，画像電子学会各会員．

柳田　益造

昭44阪大・工・電子卒．昭46同大学院修士課程了．同年NHK入局．昭53阪大大学院博士課程了．同年阪大産業科学研究所助手，昭62同助教授．昭53～54オランダ国立グローニンゲン大学音声研究所客員研究員．聴覚，音声，ディジタル信号処理の研究に従事．工博．日本音響学会，情報処理学会，IEEE各会員．

角所　収

昭25阪大・工・通信卒．昭32阪大・産業科学研究所勤務．現在，同研究所教授．工学博士．超音波，電子応用計測，医用電子装置，音声パターン認識，心理音響，ネットワーク理論，信号処理，および知的情報処理システムに関する研究に従事．1983年度Pattern Recognition Society論文賞受賞．日本音響学会，情報処理学会各会員．